

## BAB II

### TINJAUAN PUSTAKA

Penelitian dilakukan tidak terlepas dari hasil penelitian-penelitian terdahulu yang pernah dilakukan sebagai bahan perbandingan dan kajian. Adapun hasil-hasil penelitian yang dijadikan perbandingan tidak terlepas dari topik penelitian yaitu tentang menganalisis, memodelkan dan mengelompokan perilaku pengunjung (*Customer Behavior Model*) Situs Kuliah UAJY.

Pada penelitian sebelumnya. Proses analisa fungsi dimulai dengan membaca file *aces\_log*. Proses dilanjutkan dengan menganalisa baris data untuk menentukan apakah baris data tersebut mengandung string yang merepresentasikan fungsi yang sudah didefinisikan pada tahap pendefinisian simpul *graph*. Rancangan arsitektur fisik sistem menggambarkan lokasi dari komponen- komponen modul aplikasi yang digunakan untuk pemodelan tingkah laku pengunjung situs web. Langkah-langkah penelitian adalah, Memodelkan tingkah laku pengunjung, *Common Log Format* (CLF) dan identifikasi session, tahapan pemodelan tingkah laku pengunjung, rancangan arsitektur fisik sistem. (Sigit dan Kusworo,2006)

*Customer Behavior Model Graph* (CBMG), merupakan karakteristik dari suatu pola navigasi yang dilakukan suatu kelompok pengguna (pengunjung) dilihat dari sisi server, (Menasce dan Almeyda, 2000). Dengan menggunakan CBMG, maka seberapa besar beban yang dialami oleh sistem

sebagai akibat dari pola akses pengguna (penggunjung) kedalam sistem dapat diketahui. CBMG akan menangkap elemen perilaku pengguna (penggunjung) dalam hal pola navigasi, dan berguna untuk melakukan perkiraan terhadap beban yang akan dialami oleh sistem.

Model tingkah laku penggunjung suatu situs Web mengambil elemen-elemen tingkah laku penggunjung dalam arti pola navigasi, fungsi-fungsi yang digunakan, frekwensi akses terhadap fungsi-fungsi dan waktu antara pengaksesan terhadap fungsi-fungsi yang ditawarkan oleh suatu situs Web (Menasce dan Almeyda, 2000).

Pengelolaan Web server secara efektif membutuhkan *feedback*. *Feedback* ini dapat berupa informasi mengenai pengguna-pengguna yang mengakses situs Web, apa yang diakses dan statusnya, serta waktu pengaksesan. Informasi tersebut tersimpan pada log Web server terutama file access\_log. (Harol, 1997)

Daniel A. Manasce dalam artikelnya Load Testing of Web Site mengatakan bahwa developer pada umumnya mencoba melakukan pengukuran terhadap suatu aplikasi web dalam hal quality of service, throughput, dan ketersediaan (*avaliability*). *Quality of service* yang buruk akan mengakibatkan pengguna (penggunjung) frustrasi, dan ini akan berdampak pada hilangnya kesempatan usaha.

Pada umumnya, tujuan dari pengelompokan adalah untuk menyederhanakan varians dalam pengelompokan sementara memaksimalkan varians antar pengelompokan. Ini jelas bahwa jika jumlah pengelompokan dibuat

seimbang terhadap jumlah titik- titik, kita akan mencapai tujuan ini. Di sisi lain, kita ingin suatu gambaran yang rapi dari beban kerja tersebut. Sehingga, kita perlu untuk memilih suatu jumlah pengelompokan yang relative kecil semacam variansi dalam pengelompokan kecil dan variansi antar pengelompokan besar. Rasio antara variansi dalam pengelompokan dan antar pengelompokan, yang ditunjukkan  $\beta_{var}$ , dan rasio antara macam- macam koefisien dalam pengelompokan dan antar pengelompokan, yang ditunjukkan oleh  $\beta_{cv}$ , adalah penting dalam menentukan kualitas proses pengelompokan. Semakin kecil nilai dari  $\beta_{cv}$  dan  $\beta_{var}$ , maka akan semakin baik (Menasce dan Almeyda, 2000).

Aspek penting lainnya dari karakterisasi beban kerja adalah bahwa kita dapat merencanakan bentuk kunjungan yang teridentifikasi bagi tiap pengelompokan pelanggan ke dalam sumber daya perangkat lunak dan perangkat keras yang diinginkan oleh sistem tersebut (Menasce dan Almeyda, 2000).

Perilaku pengunjung situs kuliah UAJY, perlu dikelompokkan menjadi beberapa *cluster*. Salah satu metode *clustering* yang simpel, efisien, dan cepat adalah metode *k-Means*. *K-Means* adalah metode *clustering* yang mengelompokkan semua data yang kita miliki ke dalam  $k$  *cluster* yang dimana nilai  $k$  sudah ditentukan, (Arthur, 2006). serta menjadi dasar pengembangan metode clustering yang lain. (Kanungo, 2002; Bhatia, 2004; Pham, 2004; Mahdavi, 2008; Tarpey 2007).

*K-Means* merupakan salah satu metode data clustering non hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih *cluster*/kelompok. Metode ini mempartisi data ke dalam *cluster*/kelompok sehingga data yang memiliki karakteristik yang sama di kelompokkan ke dalam satu *cluster* yang sama dan data yang mempunyai karakteristik yang berbeda di kelompokkan ke dalam kelompok yang lain. Adapun tujuan dari data *clustering* ini adalah untuk meminimalisasikan *objective function* yang diset dalam proses *clustering*, yang pada umumnya berusaha meminimalisasikan variasi di dalam suatu *cluster* dan memaksimalkan variasi antar *cluster* (Agusta Y, 2007).

Pada penelitian yang dilakukan terhadap metode *k-Means clustering*, masalah yang belum diketahui adalah nilai *threshold* dalam menganalisis karakteristik pengunjung berapakah pengguna (pengunjung) yang optimal dalam menggunakan metode *k-Means* ini. Nilai *threshold* adalah nilai batasan sampai kapan iterasi (perulangan) dari *k-Means* ini diteruskan.

Untuk membendingkan penelitian pendahulu dan penelitian sekarang dapat dilihat pada tabel 2.1 dibawa ini.

Table 2.1. Perbandingan Hasil Penelitian

No	Nama Peneliti	Perbandingan Hasil Penelitian
1.	D.A Menascé; F. Almeida, A. Virgilio; Fonseca, Rodrigo, Mendes, A. Marco; 1999  Judul Penelitian : A Methodology for workload	Penelitian ini membahas tentang :  1. E-commerce workloads yang diimplementasikan dalam 2 bentuk CBMG, yaitu CBMG untuk occasional buyer, CBMG untuk heavy buyer. 2. Workload characterization methodology a) Mengidentifikasi perbedaan tiap

	<p>harakterization of E-commerce sites, to appear in the proc. of the ACM's Electronic Commerce Conference</p>	<p>session pada CBMG.</p> <ul style="list-style-type: none"> <li>b) Menghitung intensitas dari parameter beban kerja.</li> <li>c) Menetapkan sumber daya dari situs e-commerce.</li> </ul> <p>3. A CBMG based workload characterization algorithm.</p> <p>Menggunakan algoritma get sessions untuk membuat procedure get CBMG dan clustering algorithms untuk melakukan analisis clustering dan pengelompokan perilaku pengunjung situs e-commerce.</p>
2.	<p>Sigit dan Kusworo, 2006</p> <p>Judul Penelitian : Pemodelan Tingkah Laku Pengunjung Situs Web Berdasarkan Data Log Web Server</p>	<p>Penelitian ini membahas tentang :</p> <ul style="list-style-type: none"> <li>1. Membuat model tingkah laku pengunjung yang digambarkan pada CBMG.</li> <li>2. <i>Common Log Format</i> (CLF) dan Identifikasi <i>Session</i>.</li> <li>3. Tahapan pemodelan tingkah laku pengunjung, dibagi atas beberapa tahap yaitu : <ul style="list-style-type: none"> <li>a) Pendefinisian simpul graf</li> <li>b) Analisa Data Akses Web Server Log, pada analisis data akses web log server dibagi atas Penyediaan Data, Penyediaan data dilakukan dengan memodifikasi format file <code>access_log</code> untuk menentukan informasi apa saja yang dibutuhkan untuk dicatat pada file <code>access_log</code>. Penyaringan Data file multimedia (gambar, <i>icon</i>, animasi, suara dan video), <i>client-side script file</i>, dan <i>cascading style sheet file</i>. Informasi tersebut diabaikan karena merupakan bagian dari suatu <i>request</i> terhadap sebuah halaman Web.</li> <li>c) Analisa Fungsi yang digambarkan dalam flowchart.</li> </ul> </li> <li>4. Pemodelan tingkah laku, Pemodelan tingkah laku ini dilakukan berdasarkan masing-masing <i>session</i> dari pengguna.</li> <li>5. Rancangan arsitektur fisik sistem,</li> </ul>

		menggambarkan lokasi dari komponen-komponen modul aplikasi yang digunakan untuk pemodelan tingkah laku pengunjung situs Web.
3.	<p>Mateus Mas Belalawe, 2011</p> <p>Judul Penelitian : Analisis, Pemodelan Dan Pengelompokan Perilaku Pengunjung Situs Kuliah UAJY.</p>	<p>Penelitian ini membahas tentang :</p> <ol style="list-style-type: none"> <li>1. Analisis Dan Pemodelan Perilaku Pengunjung (<i>CBMG</i>). Proses analisa dimulai dengan membaca file <i>access_log</i> yang sudah terformat sesuai dengan sistem modul. Tahapan pada Analisis log situs kuliah meliputi             <ol style="list-style-type: none"> <li>a. Analisis Log Situs Kuliah Proses analisa dimulai dengan membaca file data excel atau <i>ms_access</i> yang sudah terformat sesuai dengan sistem modul, yang terdapat pada, <i>ip</i>, <i>action</i>, <i>time_normal</i>, <i>user id</i>, serta terdiri dari field berupa <i>user login</i>, <i>course view</i>, <i>user log out</i>.</li> <li>b. Pembersihan Data dimana data dari <i>access_log</i> yang akan diolah dilihat apakah merupakan data baru (berdasarkan <i>ip</i>, <i>action</i>, <i>time_normal</i>, <i>user id</i>), sehingga informasi <i>log</i> yang diabaikan berupa <i>login error</i>, <i>blog</i>, <i>forum</i>, <i>label</i>, <i>message</i>, <i>note</i>, <i>quiz</i>, <i>role</i>, <i>upload</i>, <i>survey</i>, <i>ip</i> dan <i>user id</i> yang berbeda dalam satu <i>session</i></li> <li>c. Pemodelan Tingkah Laku Model tingkah laku pengunjung ini diperoleh berdasarkan hasil ekstraksi yang dilakukan pada data log, dari pengunjung situs kuliah UAJY.</li> </ol> </li> <li>2. Pembuatan <i>Customer Visit Model</i> (CVM)</li> <li>3. Clustering Metode yang digunakan dalam penelitian tesis ini adalah <i>k-means clustering</i>. Dimana metode ini digunakan untuk mengelompokkan perilaku pengunjung situs kuliah UAJY.</li> <li>4. Menggunakan SPSS.16.0, untuk mengolah data log hasil pengunjung situs kuliah UAJY.</li> </ol>